Summary Report of the 47th Northeast Regional Stock Assessment Review Committee (SARC 47)

Members of SARC 47
John Carmichael, chair
Mike Armstrong
Kevin Stokes
Yan Jiao

Prepared for NE Region Stock Assessment Workshop
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
Woods Hole, Massachusetts

Meeting dates June 16-20, 2008
Report date July 7, 2008

Contents

Executive Summary

The 47th SARC (Stock Assessment Review Committee) met in Woods Hole, Massachusetts, from Monday, June 16, through Friday, June 20, 2008, to review the assessment of Atlantic summer flounder.

The review committee was composed of Mr. John Carmichael (South Atlantic Fishery Management Council, chair) and three scientists affiliated with the Center for Independent Experts: Dr. Kevin Stokes, Dr. Michael Armstrong, and Dr. Yan Jiao. The SARC was assisted by the NEFSC SAW Chairman, Dr. James Weinberg and his staff.

Background information and the assessment of summer flounder was presented on behalf of the Southern Demersal Working Group by Dr. Mark Terceiro, NEFSC, Dr. Mark Maunder and Mrs. Jessica Coakley. The SARC requested additional sensitivity analyses to the assessment models and reference point estimates that were accommodated by the analytical team.

The SARC concluded that the Southern Demersal Working Group successfully met all of the terms of reference. The extensive data available for the assessment appeared to be correctly compiled, and their use in the assessment and reference point analyses was in accordance with good scientific practice.

The review committee agreed that the 'ASAP' catch-age model best estimated parameters for evaluating stock status. The committee also agreed with the Southern Demersal Working Group recommended threshold reference point based on a proxy for Fmsy of F35% SPR and a target reference point based on F40%. Given current estimates of the exploitation and expected spawning stock biomass at F35% SPR, the summer flounder stock is not experiencing overfishing but is not yet rebuilt to the target biomass that is capable of producing MSY.

The review committee was impressed with level of detail devoted to presenting the summer flounder assessment data and the range of models considered. The Southern Demersal Working Group is commended for the considerable effort expended over numerous meetings to complete this assessment and address the suggestions of several prior reviews.

1 Summer Flounder Assessment Review

1.1 Introduction

Background

The 47th SARC (Stock Assessment Review Committee) met in Woods Hole, Massachusetts, from Monday, June 16, through Friday, June 20, 2008, to review the assessment of Atlantic summer flounder (*Paralichthys dentatus*).

The review committee was composed of Mr. John Carmichael (South Atlantic Fishery Management Council, chair) and three scientists affiliated with the Center for Independent Experts, University of Miami: Dr. Kevin Stokes, Dr. Michael Armstrong, and Dr. Yan Jiao. The SARC was assisted by the NEFSC SAW Chairman, Dr. James Weinberg, and his staff. The Southern Demersal Working Group was represented by Ms. Jessica Coakley, chair, and Dr. Mark Terciero.

About two weeks before the meeting, assessment documents and supporting materials were made available to the SARC via an ftp server. On the evening before the meeting, the assessment review committee met with Drs. James Weinberg and Paul Rago, NEFSC, to discuss the meeting agenda, reporting requirements, and meeting logistics. During the SARC meeting, all documents were available electronically and in print.

The meeting format opened with presentations on the Terms of Reference during which questions pertaining to the materials presented were open for question and clarification, followed by general open discussion on the Terms of Reference and concluding with dedicated, closed work sessions for the panel. The entire review committee participated in the review of each term of reference. The first 2 days and portions of the following 2 days of the meeting were open to the public and public comments were accepted during that time.

Review of Activities

The first day of the meeting (Monday afternoon) was devoted to presentations made by Dr. Mark Terceiro, on behalf of the Southern Demersal Working Group, which addressed the Terms of Reference. The committee also received a presentation on an alternative exploratory model configuration by Dr. Mark Maunder, a participant in the Southern Demersal Working Group.

In reviewing TOR 1, the assessment review committee acknowledged the high sampling intensity of the summer flounder fisheries that supported development of landings at age by market category, quarter, and primary statistical area. The committee discussed historic landings in relation to current levels and the degree of uncertainty in historical records. Improved age structure and a shift in selectivity to older ages was acknowledged. There was discussion of discard estimates, including changes over time in both discard rates and the size composition of discarded fish. The retrospective pattern observed in summer flounder stock assessments was acknowledged early in the discussions and the Committee questioned whether missing catch was a possible cause. It was reported from past studies that increasing the catch estimates in the assessment could reduce the pattern.

However, the Committee questioned whether significant missing catch was likely given the mandatory commercial reporting and low PSE's for the MRFSS estimates. Possible species misidentification (between summer and southern flounder) was questioned, to which Southern Demersal Working Group representatives indicated that it was considered unlikely to be significant.

Presentations on TOR 2 addressed the Southern Demersal Working Group's efforts to develop an integrated index from multiple surveys and explore adjustments for zero survey values. The committee questioned whether combining individual surveys would lead to some loss of information, and whether a better approach for the future could be to develop a spatially explicit model that would accommodate the spatially diverse surveys. The committee noted the efforts to explore various low values as replacements for observed zeros, but given that the approaches explored resulted in potential bias, it was suggested to consider alternatives such as error structures that would accommodate zero values or further investigation into the surveys and their CPUE calculations that might ultimately reduce zeros in final survey inputs for the assessment. Applying catch curves to survey observations is another possible approach discussed by the committee that could lead to replacement values that are potentially more biologically appropriate. The Committee noted that, although this would be useful for examining sensitivity of the assessment to treating zero observations as missing values, this is not recommended as an appropriate approach for replacing observed zero CPUE values in the final assessment.

During the presentation on TOR 3 the committee commented that a thorough fishery characterization would be beneficial to those who are not especially familiar with the summer flounder fishery. This could be accommodated through a separate document or working paper to serve as a standing reference for future reviews while avoiding excessive inflation of the assessment report. With that said, the presentations proceeded to address each portion of this TOR.

There was discussion on evaluations of regional differences in abundance and population structure and how such changes may be affected by changes in stock abundance over time. Data presented provide some evidence of a shift toward the Northern sections of the range since 1993 during increased abundance and of an increase in the size of fish found in those areas. The Committee commented that tagging studies might provide further understanding of stock movements and might provide migration information useful for spatially explicit models in the future.

Evaluation of gender differences in growth and especially mortality was acknowledged as a significant component of the assessment. The committee noted that most of the information on sex ratio and sex-specific growth rates is provided by the NEFSC survey which primarily operates offshore and may therefore only represent a portion of the stock at youngest ages. This could therefore cause bias in sex ratio estimates and maturity ogives if offshore movement of young fish is related to size- or age-influenced processes such as maturation, which is shown in the presented analyses to differ between sex. The committee discussed the need for near-shore surveys to begin collecting sex-specific biological information to further these research efforts in the future.

This led to one of the most critical aspects of the assessment, estimation of natural mortality. Discussions centered around observed maximum ages, changes in maximum

age over time, possible causes for the shift in sex ratio at age, and the various methods used to estimate natural mortality. The Committee requested likelihood profiles of alternative natural mortality estimates for the various models to be presented later. The Committee also requested a review of how the Southern Demersal Working Group arrived at their final recommended natural mortality value.

The next presentation diverged and addressed TOR 6 so that discussion of model results and configurations could be handled once the supporting issues were discussed. The working papers addressing wavelet analysis and the North Atlantic Oscillation were presented and reviewed.

The final presentations of the day addressed TOR 4, the comparison of various model approaches and their configurations and inputs. The committee initially commented that it was difficult to follow the model building process used by the Southern Demersal Working Group and requested a review of the various indicators of performance that are provided in the software and that were used by the Southern Demersal Working Group in determining the appropriate configuration. A further suggestion was made to include some measures of model performance, such as fits to indices and catch observations, or residual plots, in future reports to increase the ability of reviewers to judge work group decisions and recommendations.

Dr. Terceiro presented the modeling efforts of the Southern Demersal Working Group. The committee commented that the various alternatives all capture the significant trends in the fishery and reflect increases in biomass and decreases in exploitation over the last 10-15 years. It was also noted that the recent retrospective pattern appears to some degree in each alternative.

Dr. Mark Maunder presented his efforts to develop a sex structured model using the Stock Synthesis II software package. While the sex-structured configuration may prove useful in the future, the committee noted that the lack of sex-specific fisheries data hampers its application at this time. There was particular discussion on the estimated sex-specific natural mortality rates provided by this model and their possible support for the recommended natural mortality estimates used in the final Southern Demersal Working Group model configuration. Following these presentations the committee stood in recess for the day.

The committee opened the session Tuesday morning with a review by Dr. Terceiro of the model selection process, with particular attention given to how each survey is evaluated. Fits to various data sources and the retrospective patterns in the candidate models were reviewed in detail. This exercise proved useful to the committee, as it provided support for the Southern Demersal Working Group decisions and illuminated the model development and evaluation process. The committee felt that discussion of what was learned during each stage of development was especially helpful and that some method to capture that in future reports would increase review efficiency. Dr. Terceiro then continued with the remaining portion of the presentation on TOR 4 and TOR5, addressing the final recommended model configuration and the inclusion of data for 2007. The Committee requested greater detail on the MCMC routine included in the ASAP catch-age software.

Dr. Terceiro continued presentations on TOR 7, addressing biological reference points, and TOR 8, addressing stock projections. The committee requested further details from the Southern Demersal Working Group on the reference points required by the management process and discussed the value of limit and target references as opposed to a single reference. The committee requested additional sensitivity analyses to facilitate comparison of reference point and status changes resulting from the choice in modeling framework (and possible implicit differences), assumption and treatment of natural mortality, and the addition of new data for 2006 and 2007. There was further discussion of the retrospective pattern and associated uncertainty in stock projections.

The final presentation was made by Ms. Jessica Coakely addressing TOR 9, research recommendations. The committee appreciated the clear and concise review of past and present research needs. Various suggestions were made, which were partly incorporated into the Southern Demersal Working Group assessment report during the review. There was brief discussion of the categorization of research needs by relative importance. Evidence for commercial under-reporting as a reported major source of uncertainty was questioned. The committee discussed the need to determine the underlying cause of the retrospective pattern. The committee noted that estimation of natural mortality is the predominant and most significant major uncertainty in determining biological reference points and current stock status.

This concluded the initial presentations and the committee then moved into an additional period of general discussion and further questioning on each TOR and a reiteration of various requests for this assessment and suggestions for future assessments and assessment reports. The committee agreed that estimation of natural mortality was an important factor in the assessment and that care should be taken to examine how status changes as the model and assumed natural mortality rate are changed. Another important area of discussion at this time centered around estimation of biological reference points and what criteria should be used to select one SPR alternative over another. The committee stood in recess for the day.

The committee reconvened on Wednesday morning in a closed work session and began the task of addressing each term of reference for the report. An open session was held briefly Wednesday afternoon to review the results of the requested additional sensitivity analyses, followed by further closed sessions the remainder of Wednesday and again Thursday morning for additional writing. The committee reconvened in open session Thursday afternoon to review and edit the assessment summary report. The committee made various modifications and agreed to provide further ‘special comment’ once deliberations on all the TORs were concluded. The committee reconvened in closed session Friday morning to review report text and draft ‘special comment’ text for the summary report. The committee adjourned at 11:00 am Friday morning.

Appropriateness of the Process

The SAW/SARC process provides an appropriate means of developing and reviewing stock assessments. Overall the Stock Assessment Review Committee was impressed by the nature of the process and the efforts that went into assessment development.

The process enabled the Southern Demersal Working Group to meet several times to reach consensus on numerous critical assessment issues, including those that recently arose as well as those noted in earlier assessments. That the Working Group was supported by numerous working papers contributed by many authors representing different organizations ranging from state and federal agencies to Universities indicates the process can accommodate multiple contributions and varying viewpoints, and is not limited to work produced by a small group of people representing a single agency. While four Working Group meetings over many months may seem arduous, it is notable that no critical issues were left unaddressed due to a simple lack of time to complete those tasks charged in the Terms of Reference. Moreover, the inclusion of 2007 data into models initially developed with data through 2006 indicates that the process is very adaptable to inclusion of the most recent information, despite the need to begin resolving model input and structural details well before terminal data become available. Timeliness of information is a critical challenge facing stock assessment programs, and this assessment indicates that the SAW/SARC process can address significant issues requiring considerable time while still providing the most up-to-date results possible.

The Stock Assessment Review Committee meeting provides an effective and appropriate means of reviewing the assessment findings of the Southern Demersal Working Group. Reliance upon reviewers appointed through the Center for Independent Experts ensures that reviewers are highly qualified, truly independent, and committed to developing a thorough written report of their findings. The meeting format proved especially useful. Open session ensured public presentation and debate of the assessment, while closed working sessions enabled the panel, during the allotted workshop time, to develop a draft report requiring minimal editing in the following weeks.

1.2 SARC findings by Term of Reference

1.2.1 Characterize the Catch

> *ToR 1. Characterize the commercial and recreational catch, effort and CPUE, including descriptions of landings, discards and discard mortality.*

The Stock Assessment Review Committee considers this Term of Reference (ToR) completed successfully. Data collection programs as presented appear appropriate for estimating the quantity and the size and age composition of all significant removals due to commercial and recreational fishing. Sampling intensity has improved over time and summer flounder is recognized as one of the most intensively sampled stocks off the NE coast. The available fishery data provide a suitable basis for exploring a range of catch-at-age models to provide credible fishery management advice.

The statistical catch-at-age models implemented for the assessment using ASAP and Stock Synthesis 2 (SS2) are able to provide a close fit between estimated and observed catches at age of summer flounder, and the fitted selection patterns are in accordance with past changes in technical measures affecting selectivity. Nonetheless, all reviewed models show some retrospective pattern, and the Committee recommends continued investigation of potential catch data errors or omissions as part of a broader evaluation of

the causes of retrospective patterns in summer flounder assessments. Areas of investigation could include:

- Accuracy of recreational catch estimates and trends from the MRFSS survey and intercept program, taking into account the recent NRC review of the program;
- Appropriateness of raising procedures for estimating commercial fishery discards, and accuracy of variables such as fishing effort used in the raising procedures;
- Mortality of fish discarded by commercial and recreational fishermen, and potential changes over time.

The Committee recommends sensitivity testing of the assessment to evaluate potential errors in different aspects of fishery data. This could help target research effort toward those areas and topics where it will have greatest impact on improving the reliability of assessment models.

Male and female summer flounder have different rates of growth, maturity, and natural mortality. There is evidence for changes in population sex ratio at age which could cause bias in combined-sex assessment models and calculation of biological reference points. The Committee recommends modifying sampling schemes to allow all fishery data and both Federal and state survey indices for summer flounder to be compiled separately by sex. This will provide better understanding of sources of bias in the assessments and facilitate possible future exploration of sex-based assessments if warranted by the data. The Committee also recommends further investigation of the utility of historical sex ratio data from the NEFSC seasonal trawl surveys.

An important omission from the Southern Demersal Working Group reports to SARC 47 was an adequate characterization of the commercial and recreational fisheries. Information on the spatial distribution of fishing activities, changes in fishing effort over time, and changes in fishing gears and regulations affecting selectivity, would facilitate interpretation of the fishery data included in the assessment. The Committee recommends that future Southern Demersal Working Group reports include a suitable Working Paper on fishery characterisation.

1.2.2 Fishery Independent Surveys

ToR 2. Review methods for using fishery-independent surveys as abundance indices in assessment models.

a. Evaluate whether to combine several of the surveys into a composite survey index. If appropriate, implement this approach.

b. Develop and implement an appropriate statistical method to account for the probability of observing zeros in NEFSC survey tows.

The Committee agreed that this TOR is addressed in the SARC 47 stock assessment report. The efforts of the Southern Demersal Working Group are documented in Working Papers 3, 4, 5, and 6, and the group's findings are summarized in the assessment report.

a. *Evaluate whether to combine several of the surveys into a composite survey index. If appropriate, implement this approach.*

The Review Committee agrees with the Southern Demersal Working Group that determining how to combine separate individual surveys is a considerable research question with implications that reach beyond this single assessment. The Committee also agrees with the Working Group that the composite surveys developed may not be appropriate for this assessment. However, the intent of the ToR was addressed for this assessment in that the Southern Demersal Working Group evaluated methods of combining surveys and considered the appropriateness of such actions.

The Committee notes that differing signals from different surveys is not an unusual assessment occurrence. However, there is neither a clear benefit to combining individual sources in a simplistic manner nor a clear and efficient means to do so. Given reasonable time and resources and obvious techniques, the method used was appropriate. However, it is not unexpected that the approach did not prove especially informative given the lack of spatial overlap between the State and Federal surveys.

The survey data, even for these surveys not used in the SARC 47 stock assessment "base" model, may be valuable sources of information that provide information on spatial dynamics and could support further exploration through spatially explicit modeling that may be carried out in the future. They also provide a basis for examining spatial structure assumptions of the assessment.

b. *Develop and implement an appropriate statistical method to account for the probability of observing zeros in NEFSC survey tows.*

The Committee agreed that the evidence and analyses presented indicates it is appropriate to treat zeros in final survey catch per unit effort (CPUE) as missing values. The Southern Demersal Working Group thoroughly evaluated the suggestions and alternatives, such as using different small values, and these evaluations reveal that all methods considered for filling zero values resulted in potential bias.

The Committee recognizes that the ToR as written addresses zeros in "tows", which is slightly different than the approaches considered by the Southern Demersal Working Group and presented for review that address zeros in survey indices used as inputs to the stock assessment. However, the Committee requested and subsequently reviewed the recommendations of the 2006 S&T assessment where this recommendation originated and noted that the SAW 47 ToR referring to "tows" is slightly different than the recommendations from the 2006 "S&T" review which refers to zeros in "survey input values". Therefore, while there may be slight differences between the specific wording in the ToR and the evaluations taken by the Working Group, the Committee believes that the intent of the original SAW 47 recommendation is addressed.

The number of zero indices at low abundance indicates a high and variable proportion of zero tows. The Committee suggests that, in the future, the Southern Demersal Working Group examine the occurrence of zero observations in the individual survey tows that contribute to the overall survey CPUE at age and consider using methods such as the delta distribution approach for calculating CPUE that can accommodate zero observations in data used to develop the CPUE index. The Committee recommends a review of methods used for calculating abundance indices for the different State and

Federal surveys with a view to ensure that appropriate and consistent methods are used, including treatment of zero tows.

1.2.3 Alternative Assessment Approaches

ToR 3. Evaluate the feasibility of implementing alternative approaches to assess status of summer flounder stock and comment on any potential effects on estimates of F, SSB, and BRPs. Alternative approaches could consider:

a) Separate catch at age matrices for commercial and recreational fisheries, and resulting partial recruitment vectors for each fishery;

b) Regional differences (north, south) in catch at age matrices;

c) Potential gender differences in life span, growth rate, and natural mortality and implications of these factors for observed age- and length-specific sex ratios.

d) Strength of evidence for natural mortality rate used in the assessment; Update the estimate if appropriate.

The Committee agrees that Term of Reference 3 was completed and all of the suggested approaches were addressed with varying degrees of success. The evaluation was sufficiently comprehensive to support a final assessment model providing a credible basis for management advice.

The final catch-at-age model using ASAP represents a logical methodological advance in allowing a statistical approach to modeling errors in the fishery and survey data. The configuration of the final two-fleet ASAP model provides a bridge to past Virtual Population Analysis (VPA) configurations using ADAPT, as it places considerable weight on the catch-at-age data. Results do not differ significantly from the equivalent single-fleet ADAPT runs. However the model provides a basis to support future development and complexity that could further improve management advice for the stock. Such methods could prove particularly useful if further expansion of the stock results in differences in availability to the different fisheries, such as those in the Northeast Region and North Carolina, or if discarding practices change across regions or fisheries sectors.

a) Separate catch at age matrices for commercial and recreational fisheries, and resulting partial recruitment vectors for each fishery;

A number of multiple-fleet configurations of the ASAP and SS2 models were explored, with up to six separate catch-at-age matrices for commercial and recreational landings and discards. This has the potential for fitting time-varying selectivity patterns for different components of the fishery that have different selectivity characteristics. However, the exploration of multiple fleets suggests that the information content of historic sampling data may strongly constrain the number of catch-at-age matrices that can be modeled separately. The final assessment model chosen by Southern Demersal Working Group included only two matrices, one for all commercial and recreational landings combined, and one for all discards, regardless of source. The suggested ToR option for two separate catch at age matrices, one for all commercial catches and one for

all recreational fisheries, was not explored. The Southern Demersal Working Group approach combines retained or discarded catch-at-age data from fishing methods that have very different characteristics and are subject to sampling schemes with different error characteristics and sampling rates (e.g., port sampling, commercial observer data, MRFSS). The Committee recommends that the Southern Demersal Working Group investigate the selectivity characteristics of different fishing sectors independently of the assessment model, and the nature of sampling error for each sector, to ensure that any combined data sets are as similar as possible in terms of fishery selectivity and error structure.

b) Regional differences (north, south) in catch at age matrices;

The Southern Demersal Working Group carried out a useful descriptive exploration of spatial and temporal patterns in age compositions for the commercial landings and trawl surveys. Interpretation of the results is hindered by the lack of a detailed characterization of the fisheries. Any future developments of the assessment model to include spatial dynamics will require considerable further evaluation of spatial patterns in data, spatial dynamics of the stock, and the accuracy of positional records in the fishery data. The Southern Demersal Working Group suggested that regional differences may be addressed adequately through use of multiple fleets.

c) Potential gender differences in life span, growth rate, and natural mortality and implications of these factors for observed age- and length-specific sex ratios.

d) Strength of evidence for natural mortality rate used in the assessment; Update the estimate if appropriate.

The Southern Demersal Working Group provided a detailed evaluation of sex ratios and differences in growth and maturity between males and female summer flounder based on comprehensive data from Northeast Fisheries Science Center (NEFSC) surveys. Unfortunately, data by sex are not available from the commercial and recreational fisheries or from the various state surveys which cover areas closer inshore than the NEFSC offshore surveys. The potential magnitude of natural mortality (M) in males and females was evaluated based on published studies, comparisons with other fish stocks, and inferences from summer flounder data and model fits. The Southern Demersal Working Group recommended an increase in the combined-sex natural mortality value from 0.20 to 0.25 on the basis of their evaluations. The Committee asked for some additional ADAPT and ASAP runs to investigate the implications of the change in natural mortality for evaluation of current stock status in relation to Biological Reference Points and for the estimated yield (or TAL) associated with $F_{rebuild}$. The change in M resulted in a significant change in perception of stock status. The Committee acknowledged that the decision to change M requires careful justification and spent considerable time exploring the alternatives and their supporting evidence and also requested further detail on the process that led to the Working Group recommendation. The Southern Demersal Working Group arguments for increased M included:

- M on males is likely to be higher than on females. This is based on observed maximum ages in male and female summer flounder, declines in proportion male

with increasing age in surveys, inferences from exploratory sex-based SS2 model fits, and accumulating evidence for other fish stocks including flatfish species.

- Exploratory modeling using 'SS2' generally resulted in significantly better fits with natural mortality rates in excess of 0.2.

The Southern Demersal Working Group arrived at a weighted average M estimate of 0.25. This was calculated from values for M of 0.3 in males and 0.2 in females, inferred from recently observed maximum recorded ages of 14 years for females and 12 years for males in NEFSC surveys. A combined-sex natural mortality schedule at age was developed by assuming the initial rates by sex, an initial proportion of females at age 0 of 0.40 (based on observations from the NEFSC Fall survey), and population abundance decline over time at the sex-specific M rates. This results in slightly different M values at each age that provides a mean across ages of 0.25 for males and females combined.

The Committee noted that values of M greater than the previously adopted value of 0.20 would not be out of line with accumulating understanding of natural mortality in many other similar species worldwide. The specific M by sex will be a function of the physiological determinants of longevity in males and females as well as the abundance of predators taking different sizes of summer flounder. The specific value for summer flounder at present cannot be determined whilst fishing remains a major determinant of average longevity, and because there are no data on sex ratio in commercial and recreational catches to determine if males have different fishing mortality at age than females, though that could be possible due to the interaction of sexual dimorphism in growth and the size selectivity of the fisheries.

The Committee considered that the use of M=0.25 was an acceptable basis for the revised assessment, provided that the sensitivity of stock status evaluations to the choice of M is fully transparent. Comparative VPA (ADAPT) and catch-age (ASAP) runs at M=0.2, 0.25 and 0.33 show that all runs lead to similar estimates of current spawning stock biomass, but the B_{MSY} value declines substantially with increasing M. Hence, at M=0.25, the stock can be rebuilt by 2012 with higher $F_{rebuild}$ values and associated TALs than is the case at M=0.20. However, the long-term maximum sustainable yield (at target F_{MSY} proxy of $F_{40\%SPR}$) is lower at the higher M value as a higher proportion of the fish production is lost to natural deaths.

Some aspects of the methods applied by Southern Demersal Working Group to arrive at appropriate values of natural mortality require further consideration:

- Recent maximum ages in surveys will be strongly influenced by fishing, and also by differences in cumulative fishing mortality with increasing age in males and females;

- Sex ratios at the youngest ages could be biased in the NEFSC offshore surveys if the incidence of age-0 and age-1 fish within the survey area is dependent on migration offshore from inshore nursery areas where the sex ratio may be different. Such migrations may be related to onset of maturation (this could also lead to biased estimates of proportion mature in young flounders if a large proportion of the population of immature fish remains inshore). The Committee recommends collection of sex ratio and maturity data in inshore state surveys to

investigate this issue. More detailed spatial analysis of sex ratio at the scale of tows is warranted, particularly during autumn when spawning aggregations with skewed sex ratios may be prevalent.

1.2.4 Compare results from alternative model approaches

ToR 4. Compare results from alternative modeling approaches with those from the VPA model, to evaluate the robustness of VPA model results. Perform retrospective analyses of F, SSB, and recruitment for the models, and describe potential effects of retrospective patterns on assessment and rebuilding.

The Committee considered that this ToR had generally been addressed adequately by the Southern Demersal Working Group. The Committee noted that the ToR required development of models alternative to the ADAPT VPA to enable comparison of results and fit to various input data sources as a means of evaluating the robustness of the VPA results. However, the Southern Demersal Working Group moved beyond the requirements of the ToR by developing and recommending a revised catch-age stock assessment model drawn from one of the compared models and implemented using the ASAP package. There are some immediate consequences of this change as well as some potential long-term advantages to assessment implementation using ASAP or a similar framework (e.g., SS2). The Committee therefore asked for additional model runs to enable comparison of the effects on management-related assessment outputs. These are described below.

The models considered by the Southern Demersal Working Group included VPA assessments using ADAPT and statistical, integrated catch-age assessments implemented in either the ASAP or SS2 frameworks. All models considered rely on the same basic equations and use the same information. However, there are a number of differences, some clear and some less so, between the models even when implemented using apparently equivalent inputs and assumptions. The ADAPT models use a VPA approach and therefore solve equations backwards in time from the terminal year and treat catch-age information as exact. The ADAPT framework is restrictive in a number of ways, such as not allowing multiple fleets to be modeled separately, not supporting explicit incorporation of selectivity by fleet or time, and not including integrated estimation of reference points. Models implemented in ASAP and SS2 are fitted to data using forward projection rather than being back solved and treat catch-age and other data with error. ASAP and SS2 both allow the fitting of multiple fleets, the use of alternative error structure models and a range of alternative model forms for, for example, selectivity. In principle, very similar results should ensue if the same underlying models are fitted using any of the model frameworks. Although details might differ due to package-specific methods (for example, approach to determining start-of-series recruitment) all models should (and do) result in similar estimates of, and trends in, quantities of interest. All models as implemented therefore produce similar retrospective patterns.

Integrated approaches generally allow a more flexible approach to model fitting and have the advantage in principle of being able to produce more consistent integration of assessments, reference point calculation and projections and, importantly, better defined

estimates of uncertainty. The Committee was pleased to see adoption of these approaches for future use but was concerned to ensure that a change to any new model framework did not mask data-driven changes in assessed stock status. The Committee was also pleased to see extensive exploratory modeling undertaken in SS2 to examine the possibility of carrying out of a sex-disaggregated assessment (to account for unequal and age dependent sex ratio and differential growth) and multi-fleet modeling better to capture selectivity patterns and potentially allow advice in support of more focused management. The work revealed difficulties in fitting multiple fleets and was unlikely to succeed in dealing with sex differentiation given the underlying data availability. Although this exploratory work was not followed through to a complete candidate model configuration in this assessment, it should be continued with a view to improving understanding of the data sets and their information content and may thereby help guide future assessment modeling. It should not at this stage be considered as a viable substitute for the adopted assessment.

The Committee found it difficult to understand, from the presented working papers and assessment report, the model building and selection procedures followed by the Southern Demersal Working Group. Time was therefore allotted during the meeting for the Southern Demersal Working Group representatives to illustrate the diagnostics available during the model development process (many thanks are due in particular to Mark Terceiro for help in this respect). Following extensive review of the models considered and the process that led to final model selection, the Committee was generally satisfied with the process and decisions made. The Committee recommends that future reports contain a more transparent exposition of model specification and provide greater detail on the model building and selection procedures to facilitate review and enhance credibility.

This does not mean that a vast amount of additional material should be presented. Rather, whilst the final report should be relatively short with a text description of the process, sensitivity runs, and other necessary supporting information, a separate paper summarizing model development and selection using clear graphical presentation of diagnostics would be helpful. This is especially the case when assessments are due for review and will be useful to those who must judge the outcomes and decisions. By way of example, the main report might be very similar to the current report but written with less jargon and assuming less history with both current and past assessments. It would facilitate reading and comparisons to present the various model variants in a single reference table summarizing differences. In the separate paper it would be useful, for example, to include in standardized format the detail for major model runs of the fitted vs. observed (or residual) plots for surveys and catch-age by year, and of estimates and errors of parameters. This would allow easier understanding of what the models actually do and where fitting problems and sources of difficulty may lie. When series of models are fit, we would not propose that all such diagnostics be plotted, only those that “tell a story”, explain selection decisions, or lead to research proposals. It might be efficient as well as helpful, however, when considering a series of models, to include a tabulated summary of model fits showing likelihood components and information relevant to interpretation of the comparison. Where MCMC is used, information on priors used, posterior distributions, and MCMC performance needs to be summarized. Much or all of the information suggested is already available in the software packages used but is not necessarily useful to a reader – the ASAP interface, for example, may provide a useful

tool to the analyst in the modeling process, but the relevant information is not easily accessible to others (e.g. reviewers) in a summary way to allow comparisons between models and understand decision making related to model building. This is not an issue specific to the summer flounder assessment or to the software interface (ASAP) being used and it may be useful to consider development of standardized methods for display of outputs (e.g., using common R or similar scripts).

There is no need here to entirely revisit or repeat the model building and selection steps. Following presentations made and subsequent discussion, the Committee is satisfied that the steps taken and sensitivities considered are appropriate and that the adopted and recommended base case catch-age model implemented in ASAP is a credible basis for providing management advice. The Committee notes, however, that the adopted assessment is different in many respects to the 2006 assessment (updated data, different weight at age used in reference point calculation and terminal SSB calculations, zero survey indices treated as missing, different survey indices adopted, different selectivity pattern estimation, 2 fleets, etc). The over-riding factors influencing perceived change are the choice of M (0.25 versus 0.2), the change from using an average of the last three years of fishing mortality at age for determining partial recruitment used in reference point calculations to the use of a partial recruitment pattern derived from separately estimated selectivity for landings and discards treated as a constant since 1995, and the changed weight at age values.

Previous assessments have displayed strong retrospective patterns. The models explored all still display similar patterns, although arguably less severe in the adopted ASAP catch-age base case and in all models for the past three years. The Committee noted that even in the ASAP and SS2 implemented assessments, catch-at-age data are fitted very closely with apparent high weight placed on those data relative to survey indices. The retrospective patterns appear to be driven by the discrepancy between the well-fit catch-at-age data and the many (including the NEFSC) survey indices which imply larger biomasses at some ages in a number of recent years. Why past survey indices appear inflated relative to the signal from catch-at-age data is unclear. Potential reasons include departures from the assumption of direct proportionality between survey indices and true abundance, possibly related to the large increase in abundance and spatial extent of the stock, or a “hidden” mortality unaccounted for in the catch data (e.g., misreporting, or under-estimation of surveyed fisheries or discard losses) or due to changed environmental conditions (e.g., predation increase or environment-related range changes). The Committee is of the opinion that the underlying cause of retrospective patterns would ideally be determined and then accounted for directly in the assessment, with a consequential characterization of uncertainty in parameters and derived values therefore integrated in the model. The Committee did not think it appropriate (see ToR 8) to attempt to “fix” projections to account for retrospective pattern; the Committee therefore supports the Southern Demersal Working Group decision in this respect.

In agreeing to the base case assessment, the Committee noted that all models displayed similar trends (increasing SSB, decreasing F) but that detailed modeling decisions affect management-related outputs. The Committee was of the opinion that the differences between ADAPT, ASAP, and SS2 implemented models, and sensitivities to investigated alternative fleet structure, error structures, etc, were small compared to the differences in

outputs determined by the choice of M. Choice of M (along with weight at age and partial recruitment changes) is the over-riding issue in status determination and estimates of productivity and potential yields. The Committee was in no better a position than the Southern Demersal Working Group to choose an appropriate M value for a base case run (see ToR 3d discussion) and, as noted above, chose to agree with the Southern Demersal Working Group recommended base case but to more fully explore the dependency of management-related outputs to data updating, model assumption changes within the VPA, and use of the recommended integrated catch-age model implemented using the ASAP package (see ToR 7 discussion).

1.2.5. Estimate fishing mortality, recruitment, and biomass and characterize uncertainty.

> *ToR 5. Based on the "best" model or models, estimate fishing mortality rate, recruitment, spawning stock biomass, and total stock biomass for the current year and characterize the uncertainty of those estimates. If possible, also include estimates for earlier years with uncertainty estimates.*

The Committee agrees that this term of reference is addressed but is of the opinion that the true uncertainty could be more fully characterized The Southern Demersal Working Group put forth a preferred base case assessment and provided estimates of F, R and SSB in 2007 (the terminal year) and for the time series 1982-2007. The Southern Demersal Working Group characterized uncertainty in the estimates of F and SSB from the base case assessment for use in projection analyses (ToR 8) and for earlier years.

As with model building and selection, the Committee found it difficult to understand, from the written documentation available prior to the review, exactly how uncertainty is characterized. The Committee's understanding of the uncertainty estimates presented is that confidence intervals presented for the "earlier estimates" are derived from the ASAP catch-age base case run directly (from the Hessian matrix), but that the distributions presented for F and SSB are derived by use of a quasi-Bayesian approach utilizing MCMC and, although unspecified in the report, uniform priors. No diagnostics are provided for the MCMC. It is unclear why, if MCMC were used for the final run, it was not possible to provide MCMC-derived credible intervals for the whole time-series of estimated F and SSB, or for all derived management-related quantities. The Committee also noted (ToR 8) that using an integrated model and MCMC would in principle allow an integration of assessment and projections that would maintain coherency between all components. This approach to integrated assessment and projections is possible within ASAP but the Committee's understanding is that the method was not used due to management requirements for outputs best produced using a separate software package (AGEPRO) The Committee notes (ToR 8 below) that there are some disadvantages to the latter approach.

The Committee accepted the derived estimates and calculated uncertainties as a basis for determining stock status, making projections, and providing management advice but was generally of the opinion that better within-model uncertainty estimates should be determined and that more consideration of to the effect of major sources of uncertainty (especially M) could have been provided.

1.2.6 Examine the role of the environment on recruitment success

ToR 6. Examine and evaluate the role of the environment on past and present summer flounder recruitment success.

The Committee determined that this term of reference was completed.

Evidence presented in the two working papers (WP11, WP12) suggests that both temperature and the North Atlantic Oscillation (NAO) may correlate with summer flounder recruitment. The Committee agreed with the findings of the Southern Demersal Working Group that including environmental factors in current model configurations does not improve model performance.

1.2.7 Biological Reference Points

ToR 7. Biological Reference Points

a. Update or redefine biological reference points (BRPs; proxies for B_{MSY} and F_{MSY}), taking into account conclusions from earlier assessments and findings from TOR 6 (i.e., recruitment and the environment). Estimate uncertainty in BRPs. Comment on the scientific adequacy of existing and redefined BRPs.

b. Evaluate current stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 7a).

The Committee agreed that this Term of Reference was adequately addressed.

a. *Update or redefine biological reference points (BRPs; proxies for B_{MSY} and F_{MSY}), taking into account conclusions from earlier assessments and findings from TOR 6 (i.e., recruitment and the environment). Estimate uncertainty in BRPs. Comment on the scientific adequacy of existing and redefined BRPs.*

The Southern Demersal Working Group provided redefined estimates of biological reference points for the base case assessment. However, no measures of uncertainty in those reference points were provided. As noted above, the Committee accepts the base case assessment but emphasizes that consideration should also be given to understanding how adoption of that model, rather than a simple update of the 2006 ADAPT VPA using data from 2006-2007, affects management-related outputs. The assessment team therefore undertook additional assessments at the Committees' request and provided the results in Table 1 for consideration.

The first three columns of the table show MSY-related reference points and quantities of management interest for an assessment using the 2006 ADAPT VPA updates (S&T 2006 series). The second three columns show the same quantities for an ADAPT VPA similar to the 2006 model but using inputs updated to include 2006-2007 data (T 2007 series). The third set of three columns shows the same quantities for the Southern Demersal Working Group-adopted base case assessment using M=0.25 (middle column, labeled

F08_T2007_T2) and sensitivity runs using M of 0.2 and 0.33. Each set of three assessments is the same except that the M values used are 0.2 (as in the 2006 assessment), 0.25 (as in the adopted base case) and 0.33 (chosen given considerations discussed in section ToR 3d and because runs were available from previous Southern Demersal Working Group work).

	ADAPT VPA S&T 2006	ADAPT VPA S&T 2006	ADAPT VPA S&T 2006	ADAPT VPA T2007	ADAPT VPA T2007_M25	ADAPT VPA T2007_M33	ASAP F08_T2007_T2_M20	ASAP F08_T2007_T2	ASAP F08_T2007_T2_M33
NON-PARAMETRIC	(deterministic) M = 0.20	(deterministic) mean M=0.25	(deterministic) mean M=0.33	(stochastic) mean M=0.20	(stochastic) mean M=0.25	(stochastic) mean M=0.33	(stochastic) mean M=0.20	(stochastic) mean M=0.25	(stochastic) mean M=0.33
Fmax	0.280	0.372	0.462	0.419	0.604	1.769	0.393	0.558	1.710
MSY (mt)	21,444	19,096	17,372	14,629	13,120	10,155	16,834	12,868	10,967
SSBmax(mt)	89,411	65,606	53,650	53,384	39,314	18,489	61,653	38,547	20,973
Fterm	0.410	0.520	0.527	0.311	0.311	0.317	0.300	0.288	0.290
Yterm	13,779	13,779	13,779	10,368	10,368	10,368	10,368	10,368	10,368
SSBterm	47,498	41,449	42,441	42,142	42,919	43,711	42,185	43,363	44,066
Fterm/Fmax	1.46	1.40	1.14	0.74	0.51	0.18	0.76	0.52	0.17
Yterm/MSY	0.64	0.72	0.79	0.71	0.79	1.02	0.62	0.81	0.95
SSBterm/SSBmax	0.53	0.63	0.79	0.79	1.09	2.36	0.68	1.12	2.10
F35%	0.218	0.265	0.291	0.281	0.337	0.379	0.263	**0.310**	0.352
MSY (mt)	21,429	18,715	16,934	14,767	13,389	12,055	16,974	13,122	12,026
SSB35%(mt)	109,994	85,127	74,639	73,624	60,333	54,061	85,570	**60,074**	53,811
Fterm	0.410	0.520	0.527	0.311	0.311	0.317	0.300	0.288	0.290
Yterm	13,779	13,779	13,779	10,368	10,368	10,368	10,368	10,368	10,368
SSBterm	47,498	41,449	42,441	42,142	42,919	43,711	42,185	43,363	44,066
Fterm/F35%	1.88	1.96	1.81	1.11	0.92	0.84	1.14	0.93	0.82
Yterm/MSY	0.64	0.74	0.81	0.70	0.77	0.86	0.61	0.79	0.86
SSBterm/SSB35%	0.43	0.49	0.57	0.57	0.71	0.81	0.49	0.72	0.82
F40%	0.183	0.220	0.238	0.234	0.276	0.307	0.219	**0.255**	0.285
MSY (mt)	20,837	18,163	16,385	14,480	13,070	11,551	16,632	12,807	11,515
SSB40%(mt)	125,723	97,306	85,325	84,306	69,133	60,907	98,024	68,743	60,016
Fterm	0.410	0.520	0.527	0.311	0.311	0.317	0.300	0.288	0.290
Yterm	13,779	13,779	13,779	10,368	10,368	10,368	10,368	10,368	10,368
SSBterm	47,498	41449	42441	42,142	42,919	43,711	42,185	43,363	44,066
Fterm/F40%	2.24	2.36	2.21	1.33	1.13	1.03	1.37	1.13	1.02
Yterm/MSY	0.66	0.76	0.84	0.72	0.79	0.90	0.62	0.81	0.90
SSBterm/SSB40%	0.38	0.43	0.50	0.50	0.62	0.72	0.43	0.63	0.73

On the basis of the additional runs, the Committee makes the following comments regarding the sensitivity of estimates of biological reference points and stock status relative to the choice of model and assumed M:

- As noted by the Southern Demersal Working Group for the 2008 assessments and yield per recruit analyses, F_{MAX} is ill-defined and stock status based on comparisons with F_{MAX} and SSB at F_{MAX} are highly sensitive to changes in assumed M. The use of SPR-based reference points is therefore a more robust basis for setting thresholds.
- Consider the results for the ADAPT 2006 S&T series. When assessed and evaluated in 2007, for a choice of natural mortality of either 0.2 or 0.25, the stock would have been categorized as not yet rebuilt (i.e. $SSB_{Terminal} < SSB_{MSY}$) and subject to overfishing (i.e. $F_{Terminal}>F_{Max}$). The same conclusion would have been reached even if the threshold reference point for 2007 were based on the $F_{35\%SPR}$ estimates.
- Consider next the ADAPT 2007 T series results compared to the ASAP results. For any given M assumption, the estimates of reference points and quantities of interest are almost the same. This indicates that the change to the use of ASAP from ADAPT VPA does not cause a change in reference point values or perceptions of stock status.
- Within either of the ADAPT 2007 T series or the ASAP series, increasing the value of assumed M results in a decrease in estimated SSB_{MSY} and relatively little change in estimated $SSB_{Terminal}$. This consequently results in higher values of the ratio $SSB_{Terminal}/SSB_{MSY}$. The opposite trend occurs in $F_{Terminal}$ compared to the overfishing reference of $F_{35\%SPR}$ (or any other F reference levels). Higher assumed M's result in higher exploitation rates at each reference level, so for increased values of natural mortality, the value of the ratio $F_{Terminal}/F_{reference}$ declines.
- Therefore, unlike the ADAPT 2006 S&T series where perceived status determinations do not change over the range of M values considered, for those ADAPT VPA and ASAP catch-age assessments performed in 2008 the perception of stock status is somewhat sensitive to the choice of assumed M, particularly between the values of 0.2 and 0.25. This is because the additional data incorporated into the 2008 assessments indicate generally stable SSB and decreased exploitation when the results from 2008 are compared to those from 2006.
- Over the past two years, exploitation has continued to decline while SSB has remained fairly stable. Regardless of stock status defined by reference points, this is in line with management intentions.

The results indicate, therefore, that adoption of the ASAP catch-age base case instead of an ADAPT VPA is not the cause of change in reference points and estimates of F and SSB or status. Rather, the inclusion of new data for 2006 and 2007, and changes in weight-at-age and treatment of partial recruitment, are the determinants of status perception.

The Committee has the following additional comments:

- The Committee agrees with the Southern Demersal Working Group that the use of selectivity ogives fitted in the assessments is a better basis for partial recruitment input to the "non-parametric" reference point calculation than the previously used average of the most recent three years of estimated F at age.
- Using a short running average of weight-at-age can introduce instability to reference point estimation. The Committee recommends that future updates to biological reference points should consider methods to stabilize weight-at-age values for use in reference point calculation. The Committee notes that this recommendation applies to reference point calculation that are intended to capture overall population potential, and not necessarily to weights-at-age used for short term forecasts or current SSB calculation.
- The Committee therefore agrees that adoption of the base case ASAP catch-age assessment is a reasonable basis for providing reference point estimates and management advice.
- At a time when the stock is close to rebuilding targets and perceptions of status are critically dependent on detailed assessment choices and assumptions, it is important to provide for as much stability as possible in management and assessment processes so as not to confuse data-driven signals with changed analytical choices.

b. Evaluate current stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 7a).

The Southern Demersal Working Group provided an evaluation of stock status using reference points derived from the base case assessment. The Committee agrees with the chosen base case and the Southern Demersal Working Group's evaluation as presented.

Comparison of current assessment results with "existing biological reference points" (i.e., as agreed in 2006) is **not advised** due to the changes in assessment model and yield per recruit inputs in this assessment.

Overall and as stated by the Southern Demersal Working Group, based on the base case assessment, the stock is rebuilding towards SSB_{MSY} and F has been reduced to close to the target level. The stock is not overfished and overfishing is not occurring, although the stock is still rebuilding.

The pattern of SSB increasing towards SSB_{MSY} and F decreasing towards F_{target} can be seen in all assessment runs.

It should be remembered that the reference points and terminal F and SSB values are in fact subject to error. The Southern Demersal Working Group did not provide probability profiles of $F_{Terminal}/F_{MSY}$ or $SSB_{Terminal}/SSB_{MSY}$ for any assessments. Such information would be useful and can be derived from the MCMC assessment.

It should also be noted that the choice of M should be based on *a priori* reasoning or model fitting. Presentation of model development diagnostics should not include the derived quantities relevant to management.

1.2.8 Stock Projections

ToR 8. Stock Projections

a. Recommend what modeling approaches and data should be used for conducting single and multi-year stock projections, computing TACs or TALs, and measures of uncertainty.

b. If possible,

i. Provide numerical examples of short term projections (2-3 years) of biomass and fishing mortality rate, and characterize their uncertainty, under various TAC/F strategies and

ii. Compare projected stock status to existing rebuilding or recovery schedules, as appropriate.

This Committee agreed that this ToR is adequately addressed through the assessment report.

a. Recommend what modeling approaches and data should be used for conducting single and multi-year stock projections, computing TACs or TALs, and measures of uncertainty.

The Committee considered the AGEPRO program approach acceptable for fulfilling the ToR, but noted that although it is somewhat stochastic through the inclusion of variation in terminal year abundance and recruitment estimates, the coherent relationship among parameters within any single model run is lost using this approach. The Committee is of the opinion that the uncertainty of SSB_{MSY} is considerably underestimated by this approach. The Review Committee also advises that input assumptions and parameter choices must be consistent for both projections and reference point estimation to enable appropriate comparisons.

The review Committee recommends using an integrated approach that takes better advantage of the ASAP catch-age model's characteristics and available outputs. To better account for overall uncertainty in the assessment and key configuration decisions, the Committee recommends treating the MCMC output of parameters and variables that are required for projection simulations as consistent matrices and using the combination of parameters provided by each individual run to project the population into the future. This method would provide a coherent approach to generating management advice. Uncertainties related to individual parameters and their associations can thus be distilled into their combined influence on population projections, and thereby provide clear implications for understanding risk associated with various alternatives.

b. If possible,

i. Provide numerical examples of short term projections (2-3 years) of biomass and fishing mortality rate, and characterize their uncertainty, under various TAC/F strategies and

ii. Compare projected stock status to existing rebuilding or recovery schedules, as appropriate.

Minimal projections were provided in the assessment report and the Committee was informed that additional projections evaluating alternative scenarios will be prepared at a

later date following recommendations by the management bodies. Essentially, therefore, it is not possible for the Committee address this ToR beyond acknowledging that one-year projections are provided and that uncertainty is addressed to a degree.

As stated above, the Committee is of the opinion that true uncertainty in biomass and exploitation is likely to be higher than that represented in the available projections. The Committee suggests that future reviews could better evaluate potential rebuilding alternatives if future assessments include a set of constant catch and constant exploitation options and report the probability of rebuilding within the required timeframe and the expected year to achieve rebuilt status.

It is apparent that at this point the guiding factor in rebuilding the stock is determining the $F_{rebuild}$ and associated yield that will allow rebuilding to SSB_{MSY} by the end of 2012. Based on the figures presented, the stock could rebuild at both the proposed $F_{rebuild}$ and proposed F_{target} of $F_{35\%SPR}$ by the end of 2012.

1.2.9 Research Recommendations.

> *ToR 9. Review, evaluate and report on the status of the Research Recommendations offered in recent SARC reviewed assessments and in the 2006 "Methot" Review.*

The Committee agreed that this term of reference is adequately addressed.

The Committee appreciates the efforts of the Southern Demersal Working Group in commenting on the many previous research recommendations rather than just noting those from the last review that were addressed in this assessment. The Committee also appreciated the report format that combines all research recommendations in a single section. Two suggestions were offered by the Committee for future reports:

- Statements noting current efforts to address recommendations should note the relevant report section.
- The summary of major assessment uncertainty should be included in the ToR addressing that issue (ToR 5 in this case).

Research prioritization could be improved through appropriate simulation and sensitivity analyses constructed to identify factors which significantly affect assessment outputs relevant to management advice. Those factors that most affect advice should then be classified as high priorities. For example, the sensitivities presented in this assessment suggest that natural mortality, including differences in natural mortality between males and females, is an important issue that significantly affects assessment conclusions and management advice.

The Committee suggests the following as high priority research recommendations for summer flounder:

- Continued evaluation of natural mortality and the differences between males and females. This should include efforts to estimate natural mortality, such as through mark-recapture programs, telemetry.

- Continue efforts to improve understanding of sexually dimorphic mortality and growth patterns. This should include monitoring sex ratios and associated biological information in the fisheries and all ongoing surveys to allow development of sex-structured models in the future.
- Conduct sensitivity analyses to identify potential causes of the recent retrospective pattern. Efforts should focus on identifying factors in both survey and catch data that could contribute to the decrease in cohort abundance between initial estimates based largely on survey observations and subsequent estimates influenced by fishery dependent data as the cohort recruits to the fishery.
- Develop methods that more fully characterize uncertainty and ensure coherence between assessments, reference point calculation and projections.

2. Alternative Biological Reference Points

The SARC 47 Review Committee accepted the alternative biological reference points as recommended by the Southern Demersal Working Group and reported in the assessment (Section 1.2.7).

3. Appendices

3.1 Bibliography

Working Papers Prepared in Support of SARC 47 Terms of Referece

#	Title	Author	
1	Estimation of Commercial Fishery Discards of Summer Flounder: Update 2007 or Revise the 1989-2007 Time Series?	anon.	
2	Discard Mortality of Summer Flounder in the Inshore Trawl Fishery	Emerson Hasbrouck Tara Froehlich Kristin Gerbino John Scotti	
3	Some Approaches to the Integration of Survey Abundance Iindices used in VPA Calibration	Mark Terceiro	
4	Simulation Studies of Issues Associated with Filling Zeros in VPA Tuning Indices	Chris Legault Al Seaver	
5	Some More Thoughts on Filling Zeros in Tuning Indices: A Simple Regression Example	Chris Legault	
6	The Treatment of "Zero" Observations in the Summer Flounder ADAPT VPA Calibration	Mark Terceiro	
7	Evaluation of summer flounder life history parameters from NEFSC trawl survey data, 1992 – 2006.	Jeffrey C. Brust	
8	A Review of Natural Mortality of Summer Flounder	Rich Wong	
9	Analysis of Trends in Sex Ratio, Implications for Natural Mortality, and Variation in Age-Length Keys in Summer Flounder	Eric N. Powell Jason Morson	
10	Re-evaluation of Summer Flounder (*Paralichthys dentatus*) Stock Status Following Adjustments for Retrospective Bias and Inclusion of Trophic Effects	Victor Crecco	
11	Modeling environmental factors and summer flounder recruitment success	Mark Terceiro	
12	Wavelet Analysis of Trends in Summer Flounder YOY and Spawner-Recruit Relationships	Eric Powell	
13	Specifying Initial Conditions for Forecasting When Retrospective Pattern Present	Chris Legault and Mark Terceiro	

3.2 CIE Statement of work

Statement of Work

(NTVI Task T003 Subtask 06, prepared 19 January 2008)

External Independent Peer Review by the Center for Independent Experts

SARC 47: Summer Flounder Benchmark Stock Assessment

Meeting Date: June 16 – 20, 2008

Statement of Work (SOW) for CIE Panelists

(including a description of SARC Chairman's duties)

General

The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development (SAW Working Groups or ASMFC technical committees), assessment peer review, public presentations, and document publication.

The SARC47 review panel will be composed of three appointed reviewers from the Center of Independent Experts (CIE), and an independent chair from the South Atlantic Fishery Management Council. The panel will convene at the Woods Hole Laboratory of the Northeast Fisheries Science Center (NEFSC) in Woods Hole, Massachusetts during June 16-20, 2007 to review one assessment (Summer flounder, *Paralichthys dentatus*). In the days following the review of the assessment, the panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Overview of CIE Peer Review Process

The Office of Science and Technology implements measures to strengthen the National Marine Fisheries Service's (NMFS) Science Quality Assurance Program (SQAP) to ensure the best available high quality science for fisheries management. For this reason, the NMFS Office of Science and Technology coordinates and manages a contract for obtaining external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of stock assessments and various scientific research projects. The primary objective of the CIE peer review is to provide an impartial review, evaluation, and recommendations in accordance to the Statement of Work (SoW),

including the Terms of Reference (ToR) herein, to ensure the best available science is utilized for the National Marine Fisheries Service management decisions.

The NMFS Office of Science and Technology serves as the liaison with the NMFS Project Contact to establish the SoW which includes the expertise requirements, ToR, statement of tasks for the CIE reviewers, and description of deliverable milestones with dates. The CIE, comprised of a Coordination Team and Steering Committee, reviews the SoW to ensure it meets the CIE standards and selects the most qualified CIE reviewers according to the expertise requirements in the SoW. The CIE selection process also requires that CIE reviewers can conduct an impartial and unbiased peer review without the influence from government managers, the fishing industry, or any other interest group resulting in conflict of interest concerns. Each CIE reviewer is required by the CIE selection process to complete a Lack of Conflict of Interest Statement ensuring no advocacy or funding concerns exist that may adversely affect the perception of impartiality of the CIE peer review. The CIE reviewers conduct the peer review, often participating as a member in a panel review or as a desk review, in accordance with the ToR producing a CIE independent peer review report as a deliverable. The Office of Science and Technology serves as the COTR for the CIE contract with the responsibilities to review and approve the deliverables for compliance with the SoW and ToR. When the deliverables are approved by the COTR, the Office of Science and Technology has the responsibility for the distribution of the CIE reports to the Project Contact.

Requirements for CIE Reviewers

CIE reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models and Biological Reference Points. Expertise should include both the use of statistical catch-at-age and traditional VPA approaches. Experience with comparative studies of these approaches is especially valuable. Reviewers should also have experience in evaluating measures of model fit, identifiability, uncertainty, and forecasting. Experience with flatfish population dynamics would be useful.

Specific Activities and Responsibilities

The CIE's deliverables shall be provided according to the schedule of milestones listed on Page 6**.** The CIE reviewers, along with input and leadership from the SARC Chairman, will write the SARC Summary Report. In addition, each CIE reviewer will write an individual independent review report. These reports will provide peer-review information for a presentation to be made by NOAA Fisheries at meetings of the New England and Mid-Atlantic Fishery Management Councils in 2008. The SARC Summary Report shall be an accurate representation of the SARC panel viewpoint on how well each SAW Term of Reference was completed (please refer to Annex 1 for the SAW Terms of Reference).

The three CIE reviewers shall conduct an impartial and independent peer review in accordance with the Terms of Reference (ToR) herein. The three SARC CIE reviewers' duties shall occupy a maximum of 14 days per person (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; and several days following the open meeting to contribute to the SARC Summary Report and to produce the Independent CIE Reports).

Not covered by the CIE, the SARC chair's duties shall occupy a maximum of 15 days (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; several days following the open meeting for SARC Summary Report preparation.)

Charge to SARC panel

The panel is to determine and write down whether each Term of Reference of the SAW (see Annex 1) was or was not completed successfully during the SARC meeting. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. Where possible, the chair shall identify or facilitate agreement among the reviewers for each Term of Reference of the SAW.

If the panel rejects any of the current Biological Reference Point (BRP) proxies for B_{MSY} and F_{MSY}, the panel should explain why those particular proxies are not suitable and the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs are the best available at this time.

Roles and responsibilities

(1) Prior to the meeting

(SARC chair and CIE reviewers)
Review the reports produced by the Working Groups and read background reports.

(2) During the Open meeting

(SARC chair)
Act as chairperson, where duties include control of the meeting, coordination of presentations and discussion, making sure all Terms of Reference of the SAW are reviewed, control of document flow, and facilitation of discussion. For the

assessment, review both the Assessment Report and the Assessment Summary Report.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to discuss the stock assessment and to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(SARC CIE reviewers)
For each stock assessment, participate as a peer reviewer in panel discussions on assessment validity, results, recommendations, and conclusions. From a reviewer's point of view, determine whether each Term of Reference of the SAW was completed successfully. Terms of Reference that are completed successfully are likely to serve as a basis for providing scientific advice to management. If a reviewer considers any existing Biological Reference Point proxy to be inappropriate, the reviewer should try to recommend an alternative, should one exist.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(3) After the Open meeting

(SARC CIE reviewers)
Each reviewer shall prepare an Independent CIE Report (see Annex 2). This report should explain whether each Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified above in the "Charge to SARC panel" statement.

If any existing Biological Reference Point (BRP) proxies are considered inappropriate, the Independent CIE Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.

During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent CIE Report produced by each reviewer.

The Independent CIE Report can also be used to provide greater detail than the SARC Summary Report on specific Terms of Reference or on additional questions raised during the meeting.

(SARC chair)
The SARC chair shall prepare a document summarizing the background of the work to be conducted as part of the SARC process and summarizing whether the process was adequate to complete the Terms of Reference of the SAW. If appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the SARC Summary Report.

(SARC chair and CIE reviewers)
The SARC Chair and CIE reviewers will prepare the SARC Summary Report. Each CIE reviewer and the chair will discuss whether they hold similar views on each Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the SAW. For terms where a similar or a consensual view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

The chair's objective during this Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion.

The SARC Summary Report (please see Annex 3 for information on contents) should address whether each Term of Reference of the SAW was completed successfully. For each Term of Reference, this report should state why that Term of Reference was or was not completed successfully. The Report should also include recommendations that might improve future assessments.

If any existing Biological Reference Point (BRP) proxies are considered inappropriate, the SARC Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRP proxies are the best available at this time.

The contents of the draft SARC Summary Report will be approved by the CIE reviewers by the end of the SARC Summary Report development process. The SARC chair will complete all final editorial and formatting changes prior to

approval of the contents of the draft SARC Summary Report by the CIE reviewers. The SARC chair will then submit the approved SARC Summary Report to the NEFSC contact (i.e., SAW Chairman).

Schedule

The milestones and schedule are summarized in the table below. No later than July 7, 2008, the CIE reviewers shall submit their Independent CIE Reports to the CIE lead coordinator Mr. Manoj Shivlani via e-mail to mshivlani@ntvifederal.com and CIE regional coordinator Dr. David Sampson via e-mail to David.Sampson@oregonstate.edu.

Milestone	**Date**
CIE reviewers attend the SARC workshop to conduct peer review at Northeast Fisheries Science Center (NEFSC) in Woods Hole, MA, USA	June 16-19
SARC Chair and CIE reviewers work at the NEFSC drafting reports	June 19-20
Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair **	July 7
CIE reviewers submit Independent CIE Reports to CIE for approval	July 7
SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)	July 14
CIE provides reviewed Independent CIE Reports to NMFS COTR for approval	July 21
COTR notifies CIE of approval of reviewed Independent CIE Reports	July 28, 2008 *
COTR provides final Independent CIE Reports to NEFSC contact	July 28, 2008

* Assuming no revisions are required of the reports.
** The SARC Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the SARC chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion.

NEFSC staff and the SAW Chairman will make the final SARC Summary Report available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers, which will serve as a SAW Assessment Report.

NEFSC Contact person and SAW Chairman:
Dr. James R. Weinberg, NEFSC, Woods Hole, MA. 508-495-2352,
James.Weinberg@noaa.gov

Submission and Acceptance of CIE Reports

No later than July 21, 2008, the CIE shall provide via e-mail the final independent CIE reports and the CIE chair's summary report to the COTR William Michaels (William.Michaels@noaa.gov) at NOAA Fisheries. The COTR and alternate COTR Dr. Stephen K. Brown (Stephen.K.Brown@noaa.gov) will review the CIE reports to determine that the Term of Reference was met, notify the CIE program manager via e-mail regarding acceptance of the reports by July 28, 2008, and then distribute the reports to the NEFSC contact person.

ANNEX 1:

DRAFT Assessment Terms of Reference for SAW/SARC-47 in June, 2008

(Last Revised: Sept. 27, 2007)

Summer flounder

1. Characterize the commercial and recreational catch, effort and CPUE, including descriptions of landings, discards and discard mortality.

2. Review methods for using fishery-independent surveys as abundance indices in assessment models.

   a. Evaluate whether to combine several of the surveys into a composite survey index. If appropriate, implement this approach.

   b. Develop and implement an appropriate statistical method to account for the probability of observing zeros in NEFSC survey tows.

3. Evaluate the feasibility of implementing alternative approaches to assess status of summer flounder stock and comment on any potential effects on estimates of F, SSB, and BRPs. Alternative approaches could consider:

   a. Separate Catch at age matrices for commercial and recreational fisheries, and resulting partial recruitment vectors for each fishery.

   b. Regional differences (north, south) in catch at age matrices.

   c. Potential gender differences in life span, growth rate, and natural mortality and implications of these factors for observed age- and length-specific sex ratios.

   d. Strength of evidence for natural mortality rate used in the assessment; Update the estimate if appropriate.

4. Compare results from alternative modeling approaches with those from the VPA model, to evaluate the robustness of VPA model results. Perform retrospective analyses of F, SSB, and recruitment for the models, and describe potential effects of retrospective patterns on assessment and rebuilding.

5. Based on the "best" model or models, estimate fishing mortality rate, recruitment, spawning stock biomass, and total stock biomass for the current year and characterize the uncertainty of those estimates. If possible, also include estimates

for earlier years with uncertainty estimates.

6. Examine and evaluate the role of the environment on past and present summer flounder recruitment success.

7. Biological Reference Points

   a. Update or redefine biological reference points (BRPs; proxies for B_{MSY} and F_{MSY}), taking into account conclusions from earlier assessments and findings from TOR 6 (i.e., recruitment and the environment). Estimate uncertainty in BRPs. Comment on the scientific adequacy of existing and redefined BRPs.

   b. Evaluate current stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 7a).

8. Stock Projections

   a. Recommend what modeling approaches and data should be used for conducting single and multi-year stock projections, computing TACs or TALs, and measures of uncertainty.

   b. If possible,

      i. Provide numerical examples of short term projections (2-3 years) of biomass and fishing mortality rate, and characterize their uncertainty, under various TAC/F strategies and

      ii. Compare projected stock status to existing rebuilding or recovery schedules, as appropriate.

9. Review, evaluate and report on the status of the Research Recommendations offered in recent SARC reviewed assessments and in the 2006 “Methot” Review.

ANNEX 2: Contents of SARC CIE Independent Reports

1.

For each assessment reviewed, the report should address whether each Term of Reference of the SAW was completed successfully. For each Term of Reference, state why that Term of Reference was or was not completed successfully. To make this determination, CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable.

The report may include recommendations on how to improve future assessments.

The Independent CIE Report might also be used to provide greater detail than the SARC Summary Report on specific Terms of Reference or on additional questions raised during the meeting.

2.

If any existing Biological Reference Point (BRPs) proxies are considered inappropriate, include recommendations and justification for alternative proxies. If such alternatives cannot be identified, then indicate that the existing BRPs are the best available at this time.

3.

Any independent analyses conducted by the CIE reviewers as part of their responsibilities under this agreement should be incorporated into their Independent CIE Reports. It would also be helpful if the details of those analyses (e.g, computer programs, spreadsheets etc.) were made available to the respective assessment scientists.

4.

Additional questions that were not in the Terms of Reference but that are directly related to the assessments. This section should only be included if additional questions were raised during the SARC meeting.

5. The report shall include a list of all background material provided, a copy of the Statement of Work with Terms of Reference, and meeting agenda attached as separate appendices.

ANNEX 3: Contents of SARC Summary Report

1.

The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background, a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether each Term of Reference of the SAW was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

To make this determination, the SARC chair and CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If the CIE reviewers and SARC chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

The report may include recommendations on how to improve future assessments.

2.

If any existing Biological Reference Point (BRP) proxies are considered inappropriate, include recommendations and justification for alternative proxies. If such alternatives cannot be identified, then indicate that the existing BRPs are the best available at this time.

3.

The report shall also include the bibliography of all materials provided during SAW 47, and any papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

The report shall also include as a separate appendix the Terms of Reference used for SAW 47, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

3.3. Terms of reference

The SARC considered the summer flounder assessment in light of the terms of reference (TOR) provided to the SAW, as follows:

1. Characterize the commercial and recreational catch, effort and CPUE, including descriptions of landings, discards and discard mortality.
2. Review methods for using fishery-independent surveys as abundance indices in assessment models.
    a. Evaluate whether to combine several of the surveys into a composite survey index. If appropriate, implement this approach.
    b. Develop and implement an appropriate statistical method to account for the probability of observing zeros in NEFSC survey tows.
3. Evaluate the feasibility of implementing alternative approaches to assess status of summer flounder stock and comment on any potential effects on estimates of F, SSB, and BRPs. Alternative approaches could consider:
    a. Separate Catch at age matrices for commercial and recreational fisheries, and resulting partial recruitment vectors for each fishery.
    b. Regional differences (north, south) in catch at age matrices.
    c. Potential gender differences in life span, growth rate, and natural mortality and implications of these factors for observed age- and length-specific sex ratios.
    d. Strength of evidence for natural mortality rate used in the assessment; Update the estimate if appropriate.
4. Compare results from alternative modeling approaches with those from the VPA model, to evaluate the robustness of VPA model results. Perform retrospective analyses of F, SSB, and recruitment for the models, and describe potential effects of retrospective patterns on assessment and rebuilding.
5. Based on the "best" model or models, estimate fishing mortality rate, recruitment, spawning stock biomass, and total stock biomass for the current year and characterize the uncertainty of those estimates. If possible, also include estimates for earlier years with uncertainty estimates.
6. Examine and evaluate the role of the environment on past and present summer flounder recruitment success.
7. Biological Reference Points
    a. Update or redefine biological reference points (BRPs; proxies for B_{MSY} and F_{MSY}), taking into account conclusions from earlier assessments and findings from TOR 6 (i.e., recruitment and the environment). Estimate uncertainty in BRPs. Comment on the scientific adequacy of existing and redefined BRPs.
    b. Evaluate current stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 7a).
8. Stock Projections
    a. Recommend what modeling approaches and data should be used for conducting single and multi-year stock projections, computing TACs or TALs, and measures of uncertainty.

b. If possible,
    i. Provide numerical examples of short term projections (2-3 years) of biomass and fishing mortality rate, and characterize their uncertainty, under various TAC/F strategies and
    ii. Compare projected stock status to existing rebuilding or recovery schedules, as appropriate.

9. Review, evaluate and report on the status of the research recommendations offered in recent SARC reviewed assessments and in the 2006 “Methot” Review.